

---

DOES VOCABULARY KNOWLEDGE AFFECT LEXICAL SEGMENTATION IN  
ADVERSE CONDITIONS?

---

A thesis submitted in partial fulfilment of the  
requirements for the Degree of  
**Master of Audiology**  
in the Department of Communication Disorders  
at the University of Canterbury  
by Michelle Bishell

2015

## Table of Contents

List of Figures.....	1
List of Tables.....	2
Acknowledgements.....	3
Abstract.....	4
List of Abbreviations.....	5
1.0 Introduction.....	6
1.1 Listening in Adverse Conditions.....	6
1.2 Factors Differentiating Good Listeners from Poor Ones.....	8
1.2.1 Cognition.....	8
1.2.2 Hearing.....	12
1.2.3 Vocabulary.....	13
1.3 Lexical Segmentation.....	17
1.4 Stress as a Segmentation Cue.....	19
1.5 The Mattys, White & Melhorn (2005) Hierarchical Model of Segmentation.....	20
1.6 Vocabulary Knowledge in the Mattys et al. (2005) Model.....	22
1.7 Vocabulary Knowledge and Word Substitution Errors.....	23
1.8 Aims.....	24
1.9 Hypotheses.....	25
2.0 Method.....	26
2.1 Data Collection.....	26
2.1.1 Participants.....	26
2.1.2 Stimuli.....	26
2.1.3 Procedure.....	27
2.1.4 Data Analysis.....	28
2.1.5 Vocabulary Tests.....	28
2.2 The Current Study.....	29
2.2.1 Transcript Analysis.....	29

2.2.2 Statistical Analysis.....	30
3.0 Results.....	32
3.1 Basic Data Overview.....	32
3.2 Vocabulary and Performance.....	32
3.3 Investigating Reasons for a Relationship between Vocabulary and Performance.....	33
3.4 Relationship between Vocabulary and LBEs.....	34
3.5 Relationship between Vocabulary and WSEs.....	34
3.6 Relationship between Vocabulary and Phonemic Accuracy of WSEs.....	35
3.6.1 Consonants Correct.....	35
3.6.2 Vowels Correct.....	35
3.6.3 WSEs with No Phonemic Resemblance to the Target.....	36
3.7 Number and Proportion of Attempts.....	36
4.0 Discussion.....	37
4.1 Primary Findings.....	37
4.2 Lexical Boundary Errors.....	38
4.3 Word Substitution Errors.....	41
4.4 Number of Attempts.....	43
4.5 Limitations and Future Research.....	43
4.6 Conclusion.....	45
References.....	47

## List of Figures

Figure 1. Bar graph comparing proportion of target words correct for all participants at each signal to noise ratio.....	31
Figure 2. Scatterplot of PPVT standard score and proportion of target words correct.....	32

## List of Tables

Table 1. Summary LBE data for higher vocabulary (HV) and lower vocabulary (LV) listeners.....	33
Table 2. Summary WSE data for higher vocabulary (HV) and lower vocabulary (LV) listeners.....	34
Table 3. Average proportion of phrases with a response and average number of words per response for higher vocabulary(HV) and lower vocabulary (LV) listeners.....	35

## Acknowledgements

I would like to thank my primary supervisor, Associate Professor Megan McAuliffe, for her support, guidance and valuable advice. Thank you also to my secondary supervisor, Dr. Don Sinex, particularly for the useful feedback during the later stages of writing this thesis. To my classmates, thank you for the help and encouragement throughout the last two years. Special thanks to Sarah Kerr for her generous assistance with this thesis, especially during the data analysis stages. Many thanks also go to my friends and family for their support. Finally, to Jonathan-Adam Mukhtar, thank you for your continuous encouragement, support and patience.

## Abstract

There is significant variability in the ability of listeners to perceive degraded speech. Existing research has suggested that vocabulary knowledge is one factor that differentiates better listeners from poorer ones, though the reason for such a relationship is unclear. This study aimed to investigate whether a relationship exists between vocabulary knowledge and the type of lexical segmentation strategy listeners use in adverse conditions. This study conducted error pattern analysis using an existing dataset of 34 normal-hearing listeners (11 males, 23 females, aged 18 to 35) who participated in a speech recognition in noise task. Listeners were divided into a higher vocabulary (HV) and a lower vocabulary (LV) group based on their receptive vocabulary score on the Peabody Picture Vocabulary Test (PPVT). Lexical boundary errors (LBEs) were analysed to examine whether the groups showed differential use of syllabic strength cues for lexical segmentation. Word substitution errors (WSEs) were also analysed to examine patterns in phoneme identification. The type and number of errors were compared between the HV and LV groups. Simple linear regression showed a significant relationship between vocabulary and performance on the speech recognition task. Independent samples t-tests showed no significant differences between the HV and LV groups in Metrical Segmentation Strategy (MSS) ratio or number of LBEs. Further independent samples t-tests showed no significant differences between the WSEs produced by HV and LV groups in the degree of phonemic resemblance to the target. There was no significant difference in the proportion of target phrases to which HV and LV listeners responded. The results of this study suggest that vocabulary knowledge does not affect lexical segmentation strategy in adverse conditions. Further research is required to investigate why higher vocabulary listeners appear to perform better on speech recognition tasks.

## List of Abbreviations

BNC	British National Corpus
COCA	Corpus of Contemporary American English
DS	Delete before Strong
DW	Delete before Weak
HV	High Vocabulary
IS	Insert before Strong
IW	Insert before Weak
LBE	Lexical Boundary Error
LV	Low Vocabulary
MSS	Metrical Segmentation Strategy
NPR	No Phonemic Resemblance
PPVT	Peabody Picture Vocabulary Test
PPVT-IV	Peabody Picture Vocabulary Test fourth edition
PPVT-III-NL	Peabody Picture Vocabulary Test third edition Netherlands version
SD	Standard Deviation
SNR	Signal-to-Noise Ratio
WAIS	Wechsler Adult Intelligence Scale
WAIS-R	Wechsler Adult Intelligence Scale Revised
WSE	Word Substitution Error
WTAR	Wechsler Test of Adult Reading



## 1.0 Introduction

### 1.1 Listening in Adverse Conditions

Despite the ease with which listeners appear to understand conversational speech, the process of speech recognition is no easy task. Listeners must decipher an enormous amount of information from the incoming acoustic signal in a relatively short amount of time. The complexity of the speech recognition process is exacerbated by the suboptimal conditions in which speech frequently occurs. There are a number of ways in which the intelligibility of speech can be degraded. As a consequence, optimal listening conditions, i.e., carefully articulated clear speech produced by a healthy native speaker in a quiet environment, are an occasional occurrence rather than the norm. Mattys, Davis, Bradlow & Scott (2012) outlined three types of degradation that reduce the intelligibility of speech: (a) source degradation; (b) environmental/transmission degradation; and (c) receiver limitations. Source degradation refers to speech that has been degraded in the manner in which it was produced, resulting in reduced intelligibility compared to clear speech carefully produced by normal healthy speakers. Causes of source degradation include conversational speech and disfluencies, which decrease intelligibility by means of syllable addition, elision or reduction or by the variability in the way a single talker produces a phonetic target. For example, this variability could be because of changing register or context, or due to the coarticulation effects of the surrounding phonetic segments (Jusczyk & Luce, 2002). Accents and speech disorders also affect speech intelligibility by involving features that differ noticeably from speech norms (Clopper & Bradlow, 2008). Distinct from source degradation, environmental/transmission degradation provides other forms of distortion. Environmental/transmission degradation originates from the imperfections in the communication channel between the speaker and the listener. This situation most frequently arises from background noise or babble. However this type of distortion can also emerge from the physical environment, in the form of reverberation, or from the channel itself. For example, it may come from filtering effects on a telephone (Mattys et al., 2012). The third cause of adverse conditions is receiver limitations, which may come from impairment of either the peripheral auditory system or cognitive processing

abilities, or stem from an incomplete language model of the listener (Mattys et al., 2012). As can be seen, numerous aspects of everyday speech conditions serve to degrade the signal and make it difficult to recognise.

In relatively easy listening situations with small amounts of variability and distortion, listeners with normal hearing and speech-language abilities seem to manage with remarkable efficiency, allowing conversation to proceed relatively uninterrupted. For instance, listeners cope with the rapid pace at which conversation occurs, which may be considered a form of distortion (Mattys et al., 2012). Listeners can process an enormous amount of speech relatively quickly, reportedly accessing two to three words per second from their mental lexicon in fluent speech (Levelt, Roelofs & Meyer, 1999). Indeed, listeners can comprehend speech reasonably well even when it is presented up to twice the normal rate of speech (Orr, Friedman & Williams, 1965). In situations with low levels of degradation the majority of listeners appear to perform equally well.

However, an interesting finding of recent research is that as listening conditions degrade speech intelligibility further, individual differences in speech perception ability emerge, even among normal-hearing listeners (Gilbert, Tamati & Pisoni, 2013; Kidd, Watson & Gygi, 2007). The reasons for the differences in listener performance are unclear, although recent research has begun to investigate which individual factors may determine perceptual success. It appears likely that there may be key abilities or characteristics that allow certain listeners to glean more information from the speech signal than other listeners.

Because adverse listening conditions reflect differences in performance between listeners, they provide a useful opportunity to examine individual differences in speech perception patterns and investigate their possible causes. Analysis of listeners' error patterns may reveal information about the cues they use, and show whether good listeners use different strategies in comparison to poorer listeners. Additionally, because degraded conditions are encountered almost daily by listeners, understanding what strategies individuals use to cope with the demands of listening in adverse conditions is central to the understanding of speech perception in general.

## **1.2 Factors Differentiating Good Listeners from Poor Ones**

Listeners vary considerably in their speech recognition performance in adverse conditions (Gilbert et al., 2013; Kidd et al., 2007). Therefore, researchers in this field have aimed to uncover what sets better listeners apart from poor listeners. Existing research has investigated whether factors such as cognition, age and hearing loss contribute to the variation in listener performance. Earlier research had focused on the relationship between the state of the peripheral auditory system and listeners' ability to understand speech. However, more recently research has started to examine the role of more central processes in speech perception, including top-down influences such as cognition and vocabulary. These are discussed in more detail in subsequent sections.

### **1.2.1 Cognition**

Cognition has been hypothesised to play a role in speech perception. It is expected that listeners will utilise top-down processing resources to a greater extent when they are required to make sense of speech in the presence of background noise or ambiguity (Başkent, 2010; Warren, 1970). For example, in a complex listening environment it may be useful for the listener to suppress auditory information coming from other sources to concentrate on a single speech source. In these types of situations, it is plausible that cognitive abilities, such as working memory, could be beneficial to listeners (Schneider, Li & Daneman, 2007). In challenging listening conditions, an individual with better cognitive abilities may be better able to quickly identify what is missing from the signal and infer what may have been said. Yet the evidence for a relationship between cognition and speech perception performance has been mixed.

In the literature investigating whether there is a relationship between cognition and speech perception performance, a number of studies have found a positive relationship. In a review article of 20 studies that measured the relationship between cognition and speech recognition in noise, Akeroyd (2008) concluded that there was a link between the two, but that it was specific to particular measures of cognition and that the contribution of cognition was usually secondary to hearing acuity. The paper found that although measures of working memory were mostly effective predictors of performance,

measures of general cognitive ability, such as IQ, were not effective. Similarly, Benichov, Cox, Tun & Wingfield (2012) found that there was a relationship between understanding of speech in multitalker babble and level of cognitive ability, measured using tests of episodic memory, working memory and speed of processing. The authors reported that cognition was a predictor of listener performance, although it contributed more in the higher predictability conditions. In another study that examined whether there was a relationship between cognition and perception of sentences in multitalker babble, the findings of Tamati, Gilbert & Pisoni (2013) supported the conclusions of Akeroyd (2008). Short term and working memory scores were predictors that differentiated good from poor listeners, but executive function and nonverbal IQ scores were not. The results of Francis & Nusbaum (2009) showed further agreement with Akeroyd (2008) and Tamati et al. (2013) with regards to the effect of working memory capacity. They found that working memory was a significant factor in listener perception of low intelligibility speech and proposed that it is crucial in the formation and refinement of hypotheses of what is being said in adverse conditions. More evidence for a relationship between cognition and speech perception performance can be seen in a study examining one area of speech perception, lexical segmentation. Weiss, Gerfen & Mitchel (2010) found that when cues to segmentation in the signal were relatively equal in strength, listeners who were able to segment words from an artificial language correctly were more likely to perform well on the Simon task, a non-linguistic cognitive task that assesses the effect of conflicting cues on information processing. Moreover, in studies focusing on the elderly population, a relationship between speech recognition performance and cognition has also been observed (Humes, 2002; Jerger, Jerger & Pirozzolo, 1991).

While these studies have shown a relationship between cognitive ability and speech perception performance, other studies have found no relationship. McAuliffe, Gibson, Kerr, Anderson & LaShell (2013) found that cognition, as measured by short term and working memory tests, did not predict success in understanding dysarthric speech. Dalrymple-Alford (2014) used similar methods to test cognitive factors as McAuliffe et al. (2013) and found no relationship between scores on a short term and working memory task and performance in a perceptual task in multi-talker babble. Similarly,

Shafiro, Sheft, Risley and Gygi (2013) did not find a correlation between working memory and perception of interrupted speech. In addition, Benard, Mensink & Baskent (2014) reported that outcomes on a full-scale intelligence quotient measure did not predict participants' performance on a speech perception task involving interrupted and masked speech. Kidd et al. (2007) also used a broad scale intelligence measure. They did not find a correlation between intellectual abilities, on the basis of reported SAT scores, and auditory abilities which included speech in noise perception tasks.

To sum up, there is disagreement among these studies as to whether cognition affects listeners' speech recognition abilities. Possible contributing factors to this disagreement will be explored below.

Age is an important factor to consider when examining the role of cognition in speech perception due to the likelihood of cognitive abilities declining as age increases. Age-related cognitive decline may have partially contributed to some studies having a greater effect of cognition than others. Among the studies that found no relationship between cognition and speech perception, the majority of participants were young listeners. However, the participants in the studies that found a positive relationship were a mixture of older and younger listeners (e.g. Benichov et al., 2012; Humes, 2002; Jerger et al., 1991). Therefore, age may explain some of the differences seen in the findings. In some instances, limiting studies to younger participants may enable stronger experimental control. Subsequent studies comparing younger listeners to older listeners would provide further benefit. It is also important that studies use measures that accurately identify cognitive status when they recruit participants from a broad age range. Some studies (Benichov et al., 2012; Divenyi, Stark & Haupt, 2005) that have found a significant effect of age on speech recognition scores have attributed this effect to declines in cognitive ability, which may have not been detected by accurate cognitive testing.

In addition to participant-related variables such as age, consideration of the type of adverse conditions used to degrade the speech signal is important when exploring factors contributing to disagreement in the literature. The studies that did not find a relationship used a range of means to degrade the signal, including dysarthric speech (McAuliffe et al., 2013), broadband noise (Kidd et al.,

2007) and interrupted speech (Benard et al., 2014; Shafiro et al., 2013). The studies that did find a relationship, on the other hand, used mainly multitalker babble to reduce the intelligibility of the stimuli. Multitalker babble may require listeners to employ more cognitive resources than other types of degradation, particularly to identify the speech signal and segregate it from the competing speech sources. As Schneider et al. (2007) explain, when both the target and the masker are speech, both activate phonetic, semantic and linguistic processing systems. Therefore, interference from the competing speech may impede recognition at a cognitive level. In the other studies, where noise masking is used to distort the speech stream, it is likely only the target itself engages this level of processing, therefore involving cognitive resources to a lesser extent. Thus the nature of the adverse conditions may have a substantial impact on whether cognitive ability is a predictor variable for speech perception success. Further research using well-controlled studies could be done to investigate specific effects of different types of speech degradation on the level of cognitive involvement.

While the differences in findings among these studies may be due in part to participant age and degradation type, they do not appear to be attributable to the type of cognitive skills tested. Among the studies that did find a relationship between cognition and speech perception performance, it appears as if working memory and speed of processing measures were the best predictors for performance. Based on this pattern, it may be inferred that these types of cognitive processing are important for speech perception. However, it was also found that most of the studies that did not find a relationship also tested these types of cognitive processing, particularly working memory. Therefore, future research may need to focus on whether specific types of cognition affect speech perception in more detail to clarify their role. It is also important that the measures used accurately test the specific cognitive abilities. Some of the studies involved have used composite measures of cognition (e.g., Benichov et al., 2012), rather than investigating the individual contribution of each test measure. To isolate which aspects of cognition influenced the relationship with speech perception performance, investigation of individual parameters is needed.

In summary, although there is some evidence suggesting there is an effect of cognition on speech perception in adverse conditions, there is not complete agreement in the literature and further research needs to be done in order to draw stronger conclusions.

### **1.2.2 Hearing**

The relationship of hearing acuity to perceptual ability has been investigated thoroughly. Listeners rely on the functioning of the auditory system to extract the spectral-temporal features of the signal and send this input to the higher processing centres. Hearing loss causes a loss of audibility of what was said and, especially in the case of sensorineural hearing loss, a lack of clarity as a result of poorer frequency resolution (Bonding, 1979). As would be expected, the literature investigating the contributing factors to listener performance in adverse conditions have consistently shown hearing acuity to be a predictor variable.

Benichov et al. (2012), whose study included participants from a large age range, assessed the role of hearing acuity. Their study found that hearing accounted for partial variation in performance, particularly in conditions where stimuli were presented with low levels of context. As the level of context decreased, however, the relative contribution of hearing acuity appeared to also decrease, with cognition playing a more significant role. Further supporting evidence was found in Jerger et al. (1991), who also found that hearing acuity was the strongest predictor of performance in elderly listeners on four speech tests commonly used clinically in audiology, in comparison to cognitive status and age. Humes (2002) tested older listeners too and found a similar result. Additionally, Gordon-Salant & Fitzgibbons (1997) showed that both older and younger subjects with hearing impairment had difficulty recognising R-SPIN sentences, regardless of context or the rate of stimuli presentation, in comparison to subjects with normal hearing. These results demonstrate that hearing is significant in predicting speech perception performance irrespective of age.

In summary, the existing literature suggests that hearing acuity appears to be the predominant contributing factor to perceptual ability in older populations, though not in younger populations. It is worth noting, though, that although we know that hearing significantly affects ability to perceive

speech in noise, it is not the only factor affecting performance. Given that variability in adverse conditions exists even among normal hearing individuals, hearing loss cannot account for all of the individual differences in speech perception. In a longitudinal study, Divenyi et al. (2005) assessed a group of elderly listeners' pure-tone thresholds and their understanding of speech in quiet and in six different adverse conditions on two occasions, five years apart. After five years, auditory thresholds and speech perception performance had predictably become worse. Interestingly, though, the decline in performance on the speech perception measures was significantly faster and more variable than the decline in pure-tone thresholds. The difference in decline of the auditory performance and speech perception performance suggests that, while peripheral hearing has an important role, perception of speech also depends on other factors.

### **1.2.3 Vocabulary**

Besides cognition and hearing, vocabulary knowledge is one factor that has been hypothesised to influence how well individuals are able to identify speech in noise, though the research investigating this topic is relatively recent.

The findings of McAuliffe et al. (2013), Benard et al. (2014), and Tamati et al. (2013) all showed a positive relationship between vocabulary knowledge and speech recognition performance. McAuliffe et al. (2013) examined factors affecting the recognition and segmentation of dysarthric speech, a naturally-occurring form of degraded speech, in older and younger listeners. The listeners were asked to repeat semantically-anomalous phrases. Receptive vocabulary was measured by the Peabody Picture Vocabulary Test fourth edition (PPVT-IV) (Dunn & Dunn, 2007). In the younger group, receptive vocabulary predicted speech recognition performance in degraded conditions. Interestingly, performance on the vocabulary measures was not related to word frequency. The authors speculated that the listeners with better vocabulary knowledge utilised their greater levels of prior speech familiarity to make better use of redundancies in the acoustic signal to put together accurate lexical hypotheses. The findings of Benard et al. (2014) were in agreement with McAuliffe et al. (2013). They also examined factors affecting speech recognition, but used speech stimuli that were manipulated in a number of ways by altering speech rate, and by interrupting the signal with either



silence or filler noise. Although this study only had 12 participants, the results did show a significant correlation between listener performance and scores on the PPVT-III-NL (Dunn, Dunn & Schlichting, 2005), a measure of receptive vocabulary and verbal intelligence. The authors interpreted this correlation as evidence that verbal knowledge is a factor involved in top-down restoration mechanisms, which contribute to individual differences in understanding interrupted speech. In addition to these studies, Tamati et al. (2013) show further evidence of a relationship between vocabulary and speech perception. Their work investigated factors that differentiated listeners who performed well on a test involving speech recognition in multitalker babble, and listeners who performed poorly. They found that those who were better able to perceive speech in adverse conditions tended to have larger vocabularies, as measured by the WordFam test. The authors suggest that knowing more words and having greater lexical connectivity helped these listeners recognise words under degraded conditions. Other links between vocabulary knowledge and speech perception in degraded conditions have been noted in older adults (Janse & Adank, 2012), non-native speakers of English (Alamsaputra, Kohnert, Munson & Reichle, 2006), and in children (Munson, 2001).

Following up on this line of enquiry, recent work in our laboratory aimed to explore the effects of factors, such as vocabulary, on speech recognition in adverse conditions. Dalrymple-Alford (2014) asked listeners to repeat phrases they heard across a range of signal-to-noise (SNR) conditions. Half of the target phrases had high semantic predictability and half had low semantic predictability. The analysis of the listener responses provided further evidence that better receptive vocabulary predicts better speech recognition in adverse conditions. Specifically, the study found that this effect was clearest in the mid-range listening conditions, but not in the most highly favourable or the least favourable listening condition. Dalrymple-Alford (2014) posited that listeners with greater vocabulary knowledge utilise superior top-down processing resources to exploit intelligible glimpses in the signal, though this is yet to be systematically tested.

However, not all studies that have investigated individual differences in speech perception ability have shown a positive correlation between performance in perception tasks and vocabulary knowledge. Benichov et al. (2012) did not find a relationship between vocabulary and performance on

a perception task that asked listeners to recognise the final word of a sentence that was overlaid with masking noise. Vocabulary knowledge in this study was assessed using the WAIS vocabulary subtest (Wechsler, 1997) and WTAR (Wechsler, 2001). Both of these tests assess productive vocabulary, rather than receptive vocabulary which the other studies assessed. Jerger et al., (1991) also concluded that vocabulary knowledge did not distinguish good listeners from poor ones in their study of 200 elderly participants. Vocabulary was tested using the vocabulary subtest of the WAIS-R (Wechsler, 1981), as part of a battery of cognitive measures. The relationship between the score on this subtest and performance of the speech perception tasks was not significant.

Evidently, the conclusions of the existing research reflect some dissimilarity. The possible reasons for this disagreement will be discussed here. The cause of this disagreement may be the variety of tests used to measure vocabulary knowledge. Gilbert et al. (2013) go as far to suggest that performance on any given speech recognition measure is only an accurate representation the listener's ability in the particular conditions tested, i.e. with that particular combination of target, background competition, listener characteristics and task goal. In the studies that have explored the relationship between vocabulary and speech perception performance mentioned above, the tests used to measure vocabulary knowledge were certainly diverse in nature. For example, McAuliffe et al. (2013) and Benard et al. (2014) both used the PPVT. Tamati et al. (2013) used a word familiarity rating system, and Janse & Adank (2012) used a multiple-choice measure of receptive vocabulary. The studies that did not find a correlation, Jerger et al. (1991) and Benichov et al. (2012), both used the WAIS to assess vocabulary knowledge. These different ways of measuring vocabulary knowledge may explain some of the inconsistency in the findings between studies.

The range of participant ages among these studies is another factor that may account for some of the variation in the results. For example, Jerger et al. (1991) recruited only elderly participants, and the mean age of participants in the Benichov et al. (2012) study was 56. However, in the studies that found a relationship between vocabulary and performance, the sample populations were much younger; the mean age of the younger participants in McAuliffe et al. (2013) was 20, and in Tamati et

al. (2013) the mean age was 22 years of age. Effects of age may have produced some of the variability seen between studies.

In addition to differences in vocabulary tests and participant ages, the procedures used for measuring speech recognition ability differed across studies. For example, Benichov et al. (2012) only masked the final word of a sentence and asked the participants to identify just this final word out of the whole sentence. In contrast, in the studies of McAuliffe et al. (2013), Benard et al. (2014) and Tamati et al. (2013), listeners were required to repeat entire sentences. It is possible that the extended length of the target places extra demands on lexical knowledge, and so studies using sentences as target stimuli may be more likely to find that listeners with a large amount of lexical knowledge perform better.

In summary, only a handful of studies have investigated the effect of vocabulary knowledge on speech perception. There is some evidence that vocabulary can be a predictive factor of speech recognition performance in adverse conditions, however, the findings of two other studies did not show evidence of a relationship between vocabulary and performance. While the literature detailing this relationship is not in complete agreement, it is worth noting why there might be a link between vocabulary and speech perception in adverse conditions.

It has previously been shown that lexical factors such as word frequency, lexical neighbourhood density and lexical neighbourhood frequency affect speech perception performance (Wang & Humes, 2010; Rubenstein & Pollack, 1963; Howes, 1957; Pichora-Fuller, 2008; Luce & Pisoni, 1998). Therefore, a logical prediction is that listeners with greater lexical knowledge are better able to make use of the lexical information available in the signal. Listeners with higher vocabulary knowledge may be able to utilise a greater amount of top-down processing resources.

Allocation of top-down resources such as lexical knowledge may be particularly helpful in adverse conditions where the intelligibility of the target has been degraded. It has been hypothesised that when listening conditions are degraded, listeners exploit glimpses of more intelligible fragments of the speech signal. They use top-down processing resources to integrate this spectral and temporal

information and recover the target from the incomplete information present (Miller & Licklider, 1950; Viemeister & Wakefield, 1991; Cooke, 2006). Evidence that speech recognition in degraded conditions is mediated by top-down influences comes from fMRI studies (Sohoglu, Peele, Carlyon & Davis, 2012; Wild, Yusuf, Wilson, Peelle, Davis & Johnsrude, 2012). The hypothesis that listeners use top-down resources to put together relatively intelligible glimpses of the signal may be extended to include the effect of vocabulary knowledge. As McAuliffe et al. (2013) suggest, listeners with greater lexical knowledge may use their greater familiarity and experience with language to take advantage of the redundancies in the signal to draw accurate lexical hypotheses.

Although the existing literature suggests a link between vocabulary knowledge and listeners' ability to perceive degraded speech, the reason for such a relationship is unclear. It may be the case that higher vocabulary listeners use top-down restoration mechanisms more accurately. To date, studies have not examined whether listeners with different levels of vocabulary knowledge have different strategies for perceiving speech in difficult conditions. It remains to be seen whether listeners process speech differently, or select lexical candidates differently based on their degree of vocabulary knowledge. Analysis of one aspect of speech perception, lexical segmentation, is one way to explore why some listeners do better than others.

### **1.3 Lexical Segmentation**

Lexical segmentation is the process whereby the speech stream is broken up into word units (Jusczyk & Luce, 2002). Everyday speech arrives at the eardrum as a continuous acoustic signal, as a series of fluctuations in frequency and intensity over time. The listener must segment the incoming acoustic signal into discrete word units in order to accurately map what is being said to stored lexical representations in the mind (Dilley & McAuley, 2008). Whereas in written language the reader can rely on the spaces between words to segment a sentence, no such reliable or unambiguous boundary markers exist in spoken language (McQueen & Cutler, 2001). Examining the cues listeners use when segmenting speech can reveal differences in the strategies listeners employ to perceive speech in adverse conditions.

There is evidence that listeners use a range of cues in lexical segmentation, which is to be expected given the inherent redundancy of cues available in speech at all levels of representation from contextual cues, including semantics and pragmatics, to acoustic and prosodic cues (Church, 1987). However no cue is used exclusively or is more reliable than the others. This lack of invariant cues to segmentation leads to a large set of potential cues for the listener to consider (Weiss et al., 2010). Furthermore, the cues available may vary in the degree to which they are useful for segmentation. Listeners therefore must weight cues according to their perceived effectiveness. Sublexical cues that have been shown to assist in lexical segmentation include phonotactics (Brent & Cartwright, 1996; McQueen, 1998; van der Lugt, 2001), allophonic variation (Church, 1987), segment duration (Quené, 1993; Monaghan, White & Merks, 2013), transitional probabilities (Saffran, Newport & Aslin, 1996), prosody (Christophe, Gout, Peperkamp & Morgan, 2003) and metrical cues based on the language's rhythmic structure (Cutler & Norris, 1988; Cutler & Butterfield, 1992). These cues largely assist in identifying boundaries between words, rather than the words themselves (Newman, Sawusch & Wunnenberg, 2011).

Lexical cues also have a role in segmentation, primarily contributing to the recognition of lexical items which subsequently leads to their segmentation from the speech stream. This type of word recognition is thought to come about due to competition of multiple lexical candidates, with the successful candidate being the most acceptable parsing solution (Mattys, Melhorn & White, 2005). Vroomen & de Gelder (1995) showed that the role of lexical information in lateral inhibition during word recognition has an important function in segmentation. They suggest that lexical information has a primary role in segmentation with syllabic stress information providing supporting information, rather than vice versa. Similarly, Gow & Gordon (1995) showed the role of lexical information in segmentation was important. In their good start model (p. 352), processing at the lexical level is the driving force behind segmentation, while sublexical cues have the role of refining lexical activation by making word boundaries more prominent.

Among the numerous theoretical models of speech perception that have been proposed, there are differing points of view as to whether listeners primarily use sublexical or lexical cues to segment

speech. The various models are in agreement that speech is encoded at both sublexical and lexical levels, but differ in their accounts of where lexical segmentation occurs in the speech recognition process and how the best competitor is chosen. At the word level, some accounts have approached segmentation from a position based on multiple sublexical cues probabilistically linked with word boundaries (e.g., Christiansen, Allen & Seidenberg, 1998). In these accounts, metrical stress, phonotactic regularities, and acoustic-phonetic variants are cues that are thought to drive lexical segmentation. In other accounts, segmentation is principally the result of word recognition (e.g., Marslen-Wilson, 1987; McClelland & Elman, 1986). Critics of either approach will point out that these models have their limitations. For example, sublexical views often do not account for contradictory or incorrect sublexical cues, and, conversely, lexically driven segmentation fails when the speech signal does not map directly onto lexical representations stored in the listener's mind, meaning it is not an ideal approach in the context of learning language (Mattys et al., 2005).

## **1.4 Stress as a Segmentation Cue**

One of the sublexical cues that listeners have consistently been shown to use effectively in lexical segmentation is syllabic stress, particularly in English. The reason behind these findings is that the English language has a significant bias in its distribution of stress which gives listeners a useful indication of where lexical boundaries lie (Mattys et al., 2005). Cutler & Carter (1987) found that 90% of words in English begin with strong stress. After taking frequency into account, they claim that 75% of the strong stresses in the English language are word initial. Therefore, using stress as a cue to segmentation would appear to be an efficient listener strategy. Strong syllables are those that contain full vowel, whereas weak syllables contain a reduced vowel, usually schwa (Norris, McQueen & Cutler, 1995). The metrical segmentation strategy (MSS) hypothesises that when listening conditions are unfavourable, listeners capitalise on the high likelihood of strong syllables occurring at the beginning of words to segment speech by inserting word boundaries before strong syllables and deleting boundaries before weak syllables (Cutler & Norris, 1988; Cutler & Butterfield, 1992). Smith, Cutler, Butterfield & Nimmo-Smith (1989) showed listeners' misperceptions of speech were mostly stress-initial when listening to sentences in noise, and Cutler & Butterfield (1992) demonstrated this

same phenomenon by using faint speech as stimuli. Lexical boundary error analysis of listener transcriptions showed that listeners were more likely to mistakenly insert lexical boundaries before a strong syllable and delete boundaries after weak syllables. For example, listeners perceived the target “achieve” as “a cheap”. Moreover, computer simulations have shown stress to be a useful cue in degraded conditions as well (Harrington, Watson & Cooper, 1989).

Additional support for the MSS has been found across a number of studies involving perception of a speech target in adverse conditions. Mattys (2004) and Mattys et al. (2005) have demonstrated that stress is particularly important as listening conditions become more degraded. Speech from talkers with hypokinetic dysarthria, a naturally-occurring form of speech degradation, has also been used to elicit stress-based segmentation errors from normal listeners (Liss, Spitzer, Caviness, Adler & Edwards, 1998; Liss, Spitzer, Caviness, Adler & Edwards, 2000). Spitzer, Liss & Mattys (2007), showed listeners’ dependence on syllabic stress by manipulating of a range of segmental and suprasegmental cues in a more systematic manner, with all experimental conditions eliciting responses that conformed to the MSS. Woodfield & Akeroyd (2010) provide further evidence of stress-based segmentation. Their research investigating segmentation strategy in multitalker babble showed that young normal-hearing listeners adhered strongly to the MSS. The research showing that listeners use stress in lexical segmentation, particularly in adverse conditions, is substantial.

## **1.5 The Mattys, White & Melhorn (2005) Hierarchical Model of Segmentation**

Further literature suggests that the level of degradation of the signal is a determinant of whether stress or other cues are utilised in lexical segmentation. Based on their studies examining the relative weight listeners give to particular cues, Mattys et al. (2005) posit a hierarchical framework in which the lexical segmentation strategy employed is dictated by the particular listening conditions. Based on this model, when conditions are good, sentential context guides segmentation. As the conditions are further degraded, the strategy employed by the listener changes and the listener is reliant on lexical knowledge to facilitate segmentation. Lexically-driven segmentation is based on the

activation of lexically plausible possibilities. As listening conditions deteriorate to the point where lexical cues are unavailable, listeners tend to rely on segmental cues. Finally, when these cues are not present and listening conditions are highly degraded, stress cues are utilised most by listeners (Mattys et al., 2005). This approach reconciles the disagreement between the theoretical speech perception models focused on the bottom-up processes that primarily exploit sublexical cues and the approaches focused on the top-down processes that are driven mostly by lexical cues.

This model is supported by the findings of Sanders & Neville (2000), who investigated how listeners approach segmentation when multiple cues are present, rather than in isolation. They found that although stress was used as a segmentation cue when semantic and syntactic cues were present, its role became more prominent when these lexical cues were absent. Sanders, Neville & Woldorff (2002) carried out a similar study, but used non-native listeners. Their listeners included native Japanese and native Spanish late-learners of English, as well as near-monolingual Japanese and Spanish speakers. Their results showed that the late-learners also used semantic and lexical information to segment English speech, though they did not use syntactic information. They were able to use segmentation cues flexibly, as native speakers had done in the prior study. Both groups of late-learners used stress, and, consistent with the Mattys et al. (2005) model, used it to a greater extent when lexical and semantic information was absent. A further finding that was particularly relevant to discussion of the Mattys et al. (2005) model was that the near-monolingual listeners relied on stress to the same extent in every sentence type, regardless of whether there were lexical or semantic cues present. As listeners with nearly no English ability, they had no lexical knowledge and so were not paying attention to lexical cues at all. As a result, there was no difference to the degree to which they used stress when the lexical cues were absent. The results of Sanders et al. (2002) highlight the essence of the Mattys et al. (2005) hierarchical model: that listeners will use segmentation cues that are available in the speech stream and are usable. As higher level lexical cues become unavailable, the reliance on lower sublexical cues increases. Choe et al. (2012) also provide some support for this hierarchical model. In their analysis of listeners' transcriptions of dysarthric speech in low and high intelligibility conditions, they found that the listeners who were more successful at transcribing the



speech accurately were less reliant on syllabic stress for segmenting speech in the low intelligibility condition. The poorer listeners, in comparison, persisted with a stress-based strategy in this condition. This effect can be predicted by the Mattys et al. (2005) model in that the better listeners need not resort to using stress as a cue, as they make more effective use of the cues available at the higher tiers of the hierarchy. Further evidence for this model is present in Vitevitch & Luce (1998), who showed that lexical competition as a result of effects of neighbourhood density was key to spoken word recognition, whereas a strong facilitating effect of probabilistic phonotactics emerged only when the lexical level of cues was removed from the stimuli.

## **1.6 Vocabulary Knowledge in the Mattys et al. (2005) Model**

It is plausible that segmentation cues may be used differently by different listeners based on their individual abilities with regards to factors such as vocabulary knowledge. For example, listeners with better vocabulary knowledge may utilise cues differently to lower vocabulary listeners. If this hypothesis is correct, it may help to explain why speech recognition performance has been observed to be correlated with vocabulary knowledge. As previously mentioned Weiss et al. (2010) reported that when faced with a choice of segmentation cues with equal strength, listeners who performed better on the cognition task were more likely to select the correct cue and subsequently segment the phrase accurately. This differential use of cues depending on cognitive ability may be an effect that is also observable among listeners with different levels of vocabulary knowledge.

Using the hierarchical model described by Mattys et al. (2005), a viable hypothesis follows that listeners with a higher level of lexical ability would be less likely to need to resort to sublexical cues than listeners with a poorer vocabulary. That is, if listeners with better vocabulary knowledge are able to make more efficient use of lexical information in the perception of speech, then they may favour the lexical tier of the segmentation hierarchy.

We know from the research of Choe et al. (2012) that poorer listeners rely on stress cues for speech recognition more than better listeners in degraded listening conditions. Based on the Mattys et al. (2005) model, these results can be interpreted to suggest that better listeners are superior at

utilising lexical cues and need not resort to the stress cues as the poorer listeners do. If so, and if better listeners have superior top-down processing skills, we would expect that their lexical segmentation errors would not adhere to predictions generated by the MSS as closely as those of the poorer listeners.

## **1.7 Vocabulary Knowledge and Word Substitution Errors**

In some instances listeners accurately segment words from the speech stream by assigning the correct word boundaries, but the response produced does not accurately match the phonemic identity of the target word. Like lexical boundary errors, these word substitution errors allow observations to be made about how different listeners perceive speech. It is thought that if the acoustic stream has been correctly segmented, the resulting word substitution error should provide information about the listener's acoustic-phonetic perception of the signal (Spitzer, Liss, Caviness & Adler, 2000). Therefore, in addition to the lexical boundary information, analysis of word substitution errors may be useful in potentially reflecting differences between listeners with higher and lower levels of vocabulary knowledge.

The few studies that have investigated word substitution errors in relation to individual performance suggest that listeners who achieve higher intelligibility scores on speech perception tasks are more likely to make word substitution errors that are more phonemically similar to the target (Spitzer et al., 2000; Choe et al., 2012).

Choe et al. (2012) reported that the WSEs of the better listeners were closer to the target in their phoneme identity than those of the poorer listeners. In addition, Spitzer et al. (2000) found that listeners with higher intelligibility scores, who had familiarisation with the type of stimuli they would hear beforehand, produced WSEs that were more phonemically similar to the target than the poorer listeners, who did not have familiarisation. However, this pattern was only seen when listeners were transcribing ataxic dysarthric speech rather than hypokinetic dysarthric speech. Taking these two studies together, they suggest a trend of better listeners making WSEs that are closer in phonemic similarity to the target word, although obviously more research needs to be done in this area for firm

conclusions to be drawn. It is possible that once the word has been correctly segmented, the better performing listeners have an enhanced ability to match the acoustic-phonetic input to possible lexical candidates.

The results of the ataxic speech transcribers in Spitzer et al. (2000) provide some evidence that listeners with superior top-down knowledge can leverage their prior experience with the signal to create more accurate predictions of what was said. In their study, the transcribers of ataxic speech may have utilised their prior familiarisation with the ataxic signal to adjust their stored representations for particular phonemes. This use of top-down knowledge in phonemic mapping may be generalised to the role of vocabulary in this process. Higher vocabulary listeners may incorporate their better lexical knowledge to activate and select more phonemically accurate lexical candidates.

In summary, the literature suggests that both sublexical and lexical cues play a role in the process of lexical segmentation and word recognition. The nature of the interaction of the two appears to depend on the listening conditions and the characteristics of the individual listener. Recent studies have demonstrated that top-down effects of vocabulary knowledge may influence speech recognition ability across an array of adverse conditions. Furthermore, vocabulary knowledge may differentiate good listeners from poor listeners.

## **1.8 Aims**

Although some research investigating the effects of lexical knowledge on a listener's ability to comprehend speech in a degraded signal has been undertaken, and has provided evidence that there is a relationship between vocabulary size and perception of degraded speech, it is yet unknown why this relationship exists. The purpose of this study is to begin to address why higher vocabulary listeners might perform better than lower vocabulary listeners. A possible reason may be that listeners with better vocabulary knowledge use different strategies in lexical segmentation. The current study follows on from Dalrymple-Alford (2014), and will conduct a comprehensive lexical boundary and word substitution error analysis of this existing dataset. The aim is to determine if vocabulary size has an effect on an individual's strategy for mapping the incoming signal to the stored representations in a

listener's mind when in adverse listening conditions. In addition to the error analysis, the extent to which the listeners attempt to respond to the stimulus will also be examined to assess whether differences in speech perception performance are a result of some listeners making a greater effort to respond than others. The specific aims were as follows:

1) Determine whether higher vocabulary listeners and lower vocabulary listeners use different cue weighting, as shown using lexical boundary error analysis, when segmenting the speech stream.

2) Assess whether a relationship exists between vocabulary knowledge and number and type of word substitution errors made.

3) Investigate whether there are differences between higher and lower vocabulary listeners in the extent to which they attempt to respond to the stimulus phrases.

## **1.9 Hypotheses**

1) Listeners with larger vocabularies will pay less attention to stress cues in the lexical segmentation process than those with a smaller vocabulary. That is, it is predicted that the segmentation patterns of the listeners with better lexical knowledge will utilise their superior top-down processing and resist predictions generated by the MSS.

2) Listeners with higher vocabulary knowledge will produce WSEs that are more phonemically similar to the target phrase than those produced by lower vocabulary listeners. Higher vocabulary listeners are expected to exploit their enhanced lexical knowledge to respond with potential word candidates that have greater phonemic accuracy. Listeners with lower vocabulary knowledge, on the other hand, are hypothesised to produce WSEs that have less phonemic similarity to the target.

3) Both higher vocabulary and lower vocabulary listeners will produce a similar number of responses to the target phrases. This prediction is made on the basis that differences in performance between higher and lower vocabulary groups are due to differences in segmentation and not due to one group of listeners making more attempts than the other group.

## **2.0 Method**

The current study involved a comprehensive error pattern analysis of word recognition data previously collected in our laboratory (Dalrymple-Alford, 2014). The original study recruited young normal-hearing listeners to investigate the effects of vocabulary knowledge, age, working memory, level of semantic context and SNR on speech recognition in multitalker babble.

### **2.1 Data Collection**

All testing was completed in a single session of approximately one hour duration. Tasks were presented in the following order: (1) the Wechsler Adult Intelligence Scale (WAIS-IV) (Wechsler, 2008) working memory digit span task, (2) the Peabody Picture Vocabulary Test (PPVT-IV), (3) the WAIS-IV productive vocabulary task, and (4) the experimental listening task.

#### **2.1.1 Participants**

The data were collected from 35 participants (11 males, 24 females). Participants were 18 to 35 years old (M age=24 years). All participants were native New Zealand English speakers and had normal hearing. The majority of participants were students at the University of Canterbury, and the remainder were associates of the researcher who conducted the data collection. Pure-tone audiometry screening was carried out in a soundproof booth using supra-aural headphones. Stimuli were presented at a screening level of 15 dB HL at 500 Hz, 1 kHz, 2 kHz, 4 kHz and 8 kHz. 15 dB HL was chosen as a conservative measure of normal hearing so that hearing acuity was fully controlled for and that it would not influence listener performance. Three additional participants were excluded from the study because they exhibited thresholds of greater than 15 db HL at one or more frequencies.

The University of Canterbury Human Ethics Committee approved this study. All of the participants were informed of the procedures and aims of the study, and signed a consent form before testing began. All of the participants were compensated for the participating in the study.

#### **2.1.2 Stimuli**

The stimuli used in the perceptual task were a set of 40 six-syllable phrases. The phrases ranged from three to five words per phrase. These sentences had a mean length of 4 words (SD=0.68).

The sentences were taken from a set developed by Liss et al. (1998), modelled on sentences developed by Cutler & Butterfield (1992). These phrases had low inter-word predictability and were designed to minimise the contribution of contextual and semantic cues to intelligibility. Half the phrases had a strong-weak syllable pattern (SWSWSW), and half had a weak-strong (WSWSWS) syllable pattern. Examples of the phrases include *assume to catch control* and *admit the gear beyond*. In order to ensure that listener performance would not be affected by lack of familiarity with the test stimuli, the Range computer program was used (Nation & Heatley, 1994; Heatley, Nation & Coxhead, 2002). This program compares the stimuli against words in the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) that have been ranked according to word frequency. 94.9% of the stimulus words fell in the 4000 most common words that should be recognised by a native English speaker, so it is likely that all of the words that were used as stimuli existed in the lexicons of the participants.

The recorded stimuli were spoken by a healthy 32 year old male with a standard New Zealand accent in his normal conversational voice. The spoken phrases were recorded using an Audix HT2 Headset Condenser Microphone at a sampling rate of 44.1 kHz with 16 bits of quantization. The recorded phrases were then mixed with multi-talker babble. The stimuli were systematically degraded at four signal-to-noise ratio (SNR) conditions: -8, -4, 0 and +4. Three-talker babble was used to mask the signal. The babble was produced with overlaid utterances from three speakers from the GRID Corpus (Cooke, Barker, Cunningham & Shao, 2006).

### **2.1.3 Procedure**

The experimental task was carried out in a sound-treated booth. The phrases were recorded onto an Asus U43JC laptop computer, which was connected to a Gradson-Stadler GSI 61 two-channel audiometer during testing. The participant heard the phrases diotically through TDH-SDP supra-aural headphones. Calibration was performed prior to each test session.

The target sentences were played at 60 dB HL. The participants were asked if the volume was set at a comfortable level, and if not, was adjusted. First the listener heard four practice phrases

without background babble and was asked to verbally repeat the phrase. The participant was encouraged to guess if they were unsure what was said, and to give a partial response if they only heard part of the target phrase. When the participant showed they understood the task and could repeat the phrase without mistakes, the babble was introduced and the experimental task began.

The phrases were played using custom software made with the MATLAB program (Mathworks, Inc., 2012). This software generated four random sequences of 40 phrases. Each participant heard one of the four randomly generated sequences. Each participant heard 10 phrases at each SNR -8, -4, 0, +4. Once 10 trials at a particular SNR level had been completed, the software automatically switched to the stimuli at the next SNR level in the randomly generated sequence. Both the tester and the participant were blinded to the order of the SNR conditions.

#### **2.1.4 Data Analysis**

The responses were transcribed manually by the researcher during the session. Once testing was completed, each listener response was scored by calculating the number of words recognised accurately according to established procedures (Borrie, McAuliffe & Liss, 2012; Liss et al., 1998). Responses which differed from the target by adding or subtracting the “ed” tense ending or the plural “s”, substituted articles (“a” for “the” and vice versa), or were homonyms, were scored as correct.

#### **2.1.5 Vocabulary Tests**

Lexical knowledge of all participants was assessed during data collection. The PPVT-IV, a forced-choice measure of receptive vocabulary, was used. This study used Form A of the PPVT-IV, which has 228 items. For each item, the participants were required to select which picture, of four, best represented the lexical item spoken by the tester. The test is continued until the participant gives 8 incorrect responses in a set of 12, the ceiling set. The participant’s score is raw score is calculated by subtracting the total number of incorrect items from the number of the ceiling item. Raw scores were converted to standard scores according to the PPVT-IV manual. There were a range of receptive vocabulary scores in the group.

The Wechsler Adult Intelligible Scale fourth edition (WAIS-IV) (Wechsler, 2008) productive vocabulary test was also used to investigate the relative relationship between both productive and receptive vocabulary ability and performance on the listening task. The tester presented the participant with the test item verbally and by pointing to the word in the WAIS stimulus book at the same time. The participant was required to tell the tester what the word meant, and the response was recorded and scored according to criteria given in the WAIS manual. Raw scores were converted to standard scores using the WAIS IV manual.

Receptive vocabulary scores were moderately correlated with expressive vocabulary scores.

## **2.2 The Current Study**

In the current study, comprehensive error pattern analysis from 34 of the original 35 participants was included. The data from the participant with the median PPVT standard score was excluded from the study, so that the remaining data could be evenly be divided into two groups for later analysis. Across the entire dataset, approximately 1400 listener responses were included in the error pattern analysis.

### **2.2.1 Transcript Analysis**

The original study coded listener responses at the words correct level only. The current study conducted a detailed error analysis, with a specific focus on participants' responses for lexical boundary errors (LBEs) and word substitution errors (WSEs). The LBEs were coded as one of four error types. The error types were IS, insertion of a word boundary before a strong syllable (e.g. *indeed* becomes *the deed*); DW, deletion of a boundary before a weak syllable (e.g. *aim his* becomes *famous*); IW, insertion of a boundary before a weak syllable (e.g. *father* becomes *for the*); and DS, deletion of a boundary before a strong syllable (e.g. *or spent* becomes *suspense*). IS and DW errors are considered predicted errors according to the MSS hypothesis, and IW and DS are considered unpredicted (Cutler & Norris, 1988). This coding process also allowed for the calculation of a MSS (metrical segmentation strategy) ratio. This ratio was calculated for each participant to establish the listener's reliance on stress-based segmentation. The MSS ratio is calculated as the number of IS



errors plus the number of DW errors, divided by the total number of LBEs. An MSS ratio of greater than 0.50 indicates a stress-based approach to lexical segmentation (Spitzer et al., 2007).

In addition to LBE analysis, word substitution errors (WSEs) were examined. Analysis was based on the criteria of Spitzer et al. (2000). Word substitutions are whole words that do not violate the lexical boundaries of the intended target words. This means that the substitutions have the same number of syllables and occupy same position in the phrase as the target word, though they are phonemically different from the target in some way. These substitutions were coded according to how well they resemble the target words. They were coded as having preserved consonants, preserved vowels, or having no phonemic resemblance (NPR) to the target word, consistent with categories used by Spitzer et al. (2000). Consonant preservation was determined by the presence of the complete consonant skeleton in single syllable words, or two of the three consonants in bisyllabic words. Vowel preservation required the response to have the correct vowel in single syllable target words, and both vowels correct in bisyllabic words. WSEs were categorised as having no phonemic resemblance to the target word if neither the consonants nor the vowels were preserved.

In a small number of cases, some two-syllable responses that were considered word substitutions satisfied the criteria for both the consonants correct and the vowels correct categories. For example, in one instance the target word was “sinking” and the participant response was “thinking”. The requirements for vowel preservation were satisfied: both the vowels were correct. The criterion for consonant preservation is also satisfied because two of the three consonants are correct. Therefore, this word substitution error was coded as having both vowels and consonants correct. In these situations, the response was counted as having two word substitution errors. This was done to ensure that during the later analysis the categories of consonants correct and vowels correct would remain exclusive from each other.

### 2.2.2 Statistical Analysis

First the extent of the relationship between vocabulary and performance on the perceptual task was determined using linear regression analysis. The participants were then split into two groups: a higher vocabulary (HV) and a lower vocabulary (LV) group.

Independent samples t-tests were used to examine differences between the vocabulary groups regarding lexical boundary error patterns, including the number of LBEs they made and their average MSS ratios. Differences in WSEs made by the higher and lower vocabulary listeners were also examined by performing independent samples t-tests on the proportions of errors with consonants correct, vowels correct, and no phonemic resemblance. Significant differences were regarded as a rejection of the null hypothesis.

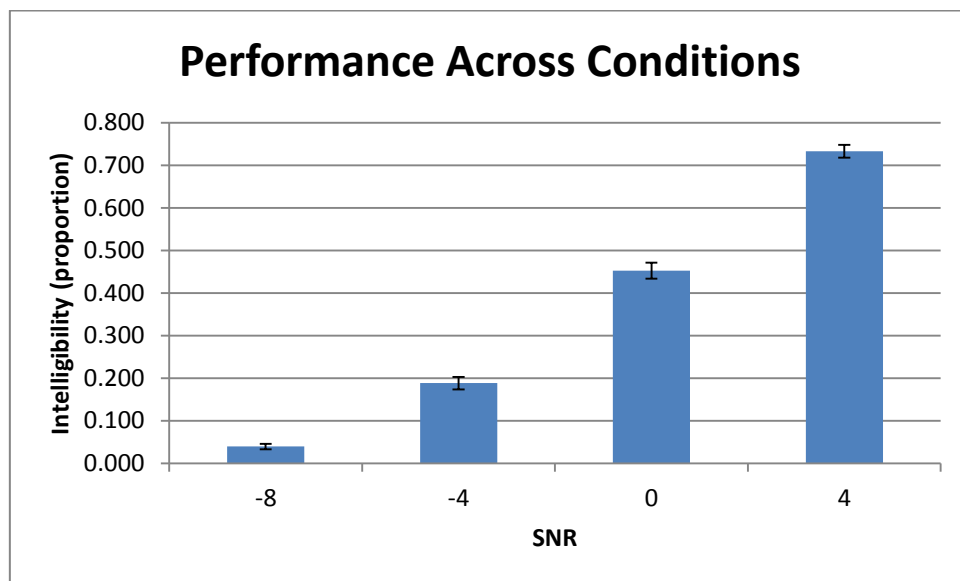
Independent samples t-tests were used to determine whether there were significant differences between higher and lower vocabulary listeners in terms of the proportion of target phrases to which they responded. This comparison was done in order to examine whether differences in performance could be attributed to some listeners making more attempts to respond than others.

It was originally intended that lexical boundary and word substitution errors would be examined at the four individual SNR conditions to investigate whether error patterns differed according to the level of intelligibility. However, because the number of errors produced by the listeners was smaller than expected, particularly in the -8 and +4 SNR conditions, these results are presented collapsed across SNR.

## 3.0 Results

### 3.1 Basic Data Overview

Listener intelligibility scores improved as SNR increased. Figure 1 shows the mean intelligibility scores as a proportion of the target words correct across the four SNR conditions. Speech recognition performance ranged from 3.9% in the -8 SNR condition to 73.2% in the +4 SNR condition.



**Figure 1:** Bar graph comparing proportion of target words correct for all participants at each signal-to-noise ratio (SNR). Error bars represent the standard error of the mean.

### 3.2 Vocabulary and Performance

The mean PPVT standard score for all listeners was 109.5 (SD=10.6), which was slightly higher than the normative mean (100). Therefore, as a group, the participants had slightly better receptive vocabulary knowledge than the general population. Figure 2 displays the association between listeners' PPVT standard score and their mean intelligibility score. Each datapoint represents a single participant.

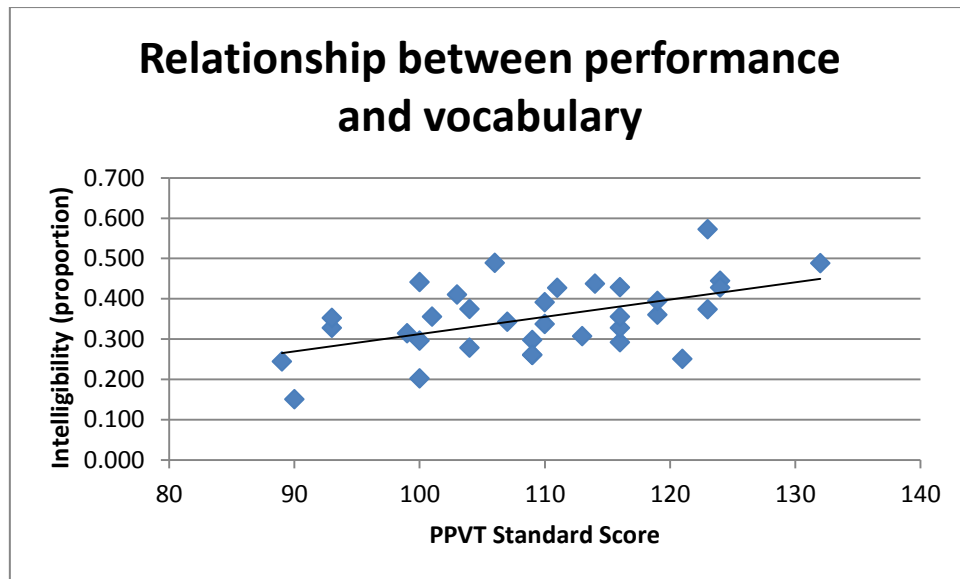


Figure 2: Scatterplot of PPVT standard score and proportion of target words correct.

A linear regression was performed. The primary aim of investigating this correlation was to examine whether there was a relationship between listeners' level of vocabulary knowledge and their ability to comprehend speech in adverse conditions. Vocabulary knowledge significantly predicted intelligibility scores,  $\beta=0.004$ ,  $t(32)=3.42$ ,  $p < 0.001$ . Vocabulary also explained a significant proportion of variance in intelligibility scores,  $R^2 = 0.27$ ,  $F(1, 32) = 11.67$ ,  $p = <0.001$ . Receptive vocabulary therefore contributed significantly to variation on the speech recognition task, with participants with higher PPVT scores recognising a greater proportion of the target phrases.

### 3.3 Investigating Reasons for a Relationship between Vocabulary and Performance

In order to investigate why there is a relationship between vocabulary and performance, the participants were divided into two groups, a higher vocabulary (HV) group and a lower vocabulary (LV) group based on a median split of the participants' standardised PPVT scores. The HV group had a mean PPVT standard score of 118 ( $SD=5.9$ ), and the LV group had a mean PPVT standard score of 101 ( $SD=6.5$ ). A comparison of vocabulary group and performance shows that the mean proportion of the target speech that listeners recognised correctly was 0.389 ( $SD=0.08$ ) in the HV group and 0.317 ( $SD=0.09$ ) in the LV group.

### 3.4 Relationship between Vocabulary and LBEs

The number and type of LBEs were examined to identify whether there were differences in lexical segmentation strategy between the two vocabulary groups. Summary data is provided in Table 1 below.

**Table 1: Summary LBE data for higher vocabulary (HV) and lower vocabulary (LV) listeners.**

	<b>HV</b>	<b>LV</b>
<b>Total no. of LBEs</b>	61	80
<b>Average MSS ratio</b>	0.75 (0.26)	0.67 (0.36)
<b>Proportion IS</b>	0.51	0.49
<b>Proportion DW</b>	0.30	0.30
<b>Proportion IW</b>	0.07	0.11
<b>Proportion DS</b>	0.13	0.10

An independent samples t-test showed no significant difference between the mean number of LBEs for the HV group ( $M=3.59$ ,  $SD=2.43$ ) and LV group ( $M=4.71$ ,  $SD=2.82$ );  $t(32)=-1.238$ ,  $p=0.225$ .

A further independent samples t-test was performed comparing the mean MSS ratios of participants in the HV and LV groups. There was no significant difference between the mean MSS ratio of the HV group ( $M=0.75$ ,  $SD=0.26$ ) and LV group ( $M=0.67$ ,  $SD=0.37$ );  $t(32)=0.775$ ,  $p=0.444$ .

### 3.5 Relationship between Vocabulary and WSEs

Summary data describing the WSE performance of higher and lower vocabulary listeners is provided in Table 2 below. As can be seen, the greatest proportion of WSEs across the two groups had no phonemic resemblance to the target word.

**Table 2: Summary WSE data for higher vocabulary (HV) and lower vocabulary (LV) listeners.**

	<b>HV</b>	<b>LV</b>
<b>Total no. of WSEs</b>	271	262
<b>Ave no. of WSEs</b>	15.9 (5.38)	15.4 (5.28)
<b>Average proportion WSEs with consonants correct</b>	0.22 (0.09)	0.24 (0.15)
<b>Average proportion WSEs with vowels correct</b>	0.33 (0.15)	0.27 (0.14)
<b>Average proportion WSEs with no phonemic resemblance</b>	0.46 (0.13)	0.49 (0.21)

In order to compare the mean number of WSEs between the HV and LV groups an independent samples t-test was performed. There was no significant difference in the number of WSEs between the HV group ( $M=15.9$ ,  $SD=5.37$ ) and the LV group ( $M=15.4$ ,  $SD=5.28$ );  $t(32)=0.290$ ,  $p=0.774$ .

### **3.6 Relationship between Vocabulary and Phonemic Accuracy of WSEs**

#### **3.6.1 Consonants Correct**

An independent samples t-test was conducted to compare the mean proportion of WSEs that were judged to have consonants correct between the two vocabulary groups. There was no significant difference between the HV group ( $M=0.217$ ,  $SD=0.089$ ) and the LV group ( $M=0.236$ ,  $SD=0.155$ );  $t(32)=-0.456$ ,  $p=0.652$ .

#### **3.6.2 Vowels Correct**

An independent samples t-test was carried out to compare the mean proportion of WSEs that were considered to have vowels correct between the two vocabulary groups. No significant difference was found between the HV group ( $M=0.328$ ,  $SD=0.147$ ) and the LV group ( $M=0.275$ ,  $SD=0.139$ );  $t(32)=1.093$ ,  $p=0.283$ .

### 3.6.3 WSEs with No Phonemic Resemblance to the Target

In order to compare the mean proportion of WSEs that had no phonemic resemblance to the target between the two vocabulary groups, an independent samples t-test was performed. There was no significant difference between the HV group ( $M=0.455$ ,  $SD=0.129$ ) and the LV group ( $M=0.489$ ,  $SD=0.213$ );  $t(32)=-0.560$ ,  $p=0.579$ .

### 3.7 Number and Proportion of Attempts

The number of responses given by listeners in each of the two vocabulary groups was examined in order to investigate whether the HV group's higher speech recognition scores could be attributed to the listeners in the HV group making more response attempts than those in the LV group. Table 3 shows the mean proportion of the total number of target phrases to which participants responded, and the mean number of words in the responses given.

**Table 3: Average proportion of phrases with a response and average number of words per response for higher vocabulary (HV) and lower vocabulary (LV) listeners.**

	<b>HV</b>	<b>LV</b>
<b>Average proportion of phrases with a response</b>	0.74 (0.09)	0.71 (0.13)
<b>Average no. of words in each response</b>	2.26 (0.47)	2.06 (0.45)

An independent-samples t-test was conducted to compare the mean proportion of phrases with a response between the higher vocabulary and the lower vocabulary groups. There was no significant difference in proportions for the HV group ( $M=0.74$ ,  $SD=0.09$ ) and LV group ( $M=0.71$ ,  $SD=0.13$ );  $t(32)=0.755$ ,  $p=0.455$ .

## 4.0 Discussion

Young healthy listeners with normal hearing exhibit variation in their speech recognition performance in adverse listening conditions (Gilbert et al., 2013; Kidd et al., 2007). Prior research has shown that listeners with superior vocabulary knowledge perform better in these speech recognition tasks (McAuliffe et al., 2013; Benard et al., 2014; Tamati et al., 2013; Janse & Adank, 2012; Alamsaputra et al., 2006; Munson, 2001; Dalrymple-Alford, 2014). The current study hypothesised that higher vocabulary listeners may exhibit differential cue use in lexical segmentation. Thus, the aim of the current study was to investigate whether an individual's level of vocabulary knowledge had an effect on the types of cues they use in the segmentation and identification of words in speech.

The hypotheses of this study were that listeners with a higher level of vocabulary knowledge would display different error patterns to listeners with a lower level of vocabulary knowledge as a result of differences in cue use to uncover what was said. It was expected that listeners with better vocabulary knowledge would give less weighting to syllabic stress as a cue to segmentation than the listeners with poorer vocabulary because the higher vocabulary listeners would attend to lexical cues more. It was also predicted that the word substitution errors produced by the higher vocabulary listeners would have more phonemic resemblance to the target word than those produced by the lower vocabulary listeners, because higher vocabulary listeners were expected to use their greater lexical knowledge upon hearing the acoustic signal to produce more accurate predictions of what was said in the target phrase. The 34 participants were divided evenly into a higher vocabulary and a lower vocabulary group based on their score on the PPVT receptive vocabulary test.

### 4.1 Primary Findings

The study reported 3 primary findings: (a) Lexical boundary error analysis revealed no significant differences between vocabulary groups in the extent to which listeners used stress as a cue to segmentation; (b) word substitution error analysis showed that there was no difference in the proportion of word substitution errors that were phonemically similar to the target between the errors produced by the higher vocabulary listeners and those made by the lower vocabulary listeners; (c)



higher and lower vocabulary listeners did not differ in the proportion of the target phrases to which they provided a response nor in the number of lexical boundary or word substitution errors they made.

## **4.2 Lexical Boundary Errors**

Listeners from both the higher and lower vocabulary groups exhibited similar lexical segmentation patterns. The MSS ratios show that both groups consistently relied on stress to segment the speech signal. Both groups had MSS ratios above 0.5, and therefore were more likely to insert a boundary before a strong syllable and delete a boundary before a weak syllable. Approximately half of all LBEs from both higher and lower vocabulary listeners involved the insertion of a word boundary before a stressed syllable. These findings provide further evidence to support the MSS hypothesis (Cutler & Butterfield, 1992; Cutler & Norris, 1988). Therefore, the current study adds to the already large evidence base that highlights the MSS as a particularly robust strategy for segmenting speech.

However, it was predicted prior to this investigation that only the lower vocabulary listeners would show such extensive reliance on syllabic stress. The hypothesis of this study regarding lexical segmentation was drawn from the hierarchical framework of segmentation described in Mattys et al. (2005). The current study findings did not support this hypothesis and were also inconsistent with those of Choe et al. (2012), where poorer listeners relied on syllabic stress more than better listeners when listening to low intelligibility degraded speech. In the present study, the listeners who had performed more poorly on the speech perception task in the previous study (i.e. those with lower vocabulary knowledge) and the listeners who performed better (who had higher vocabulary) used stress to the same extent when listening to speech in adverse conditions. Exploration of the possible reasons for the current findings may address the discrepancy in results between this and prior studies.

One explanation for the lexical segmentation results is that all listeners attend to cues to the same extent during segmentation, regardless of vocabulary knowledge. That is, the difference in overall performance between vocabulary groups is not due to differences in the way the two groups approach segmentation. Both groups of listeners used stress in the listening conditions of the current

study, and, following Mattys et al. (2005)'s approach, if more lexical information had been present both groups would have used stress to a lesser extent in response. The vocabulary differences between listeners may have influenced their performance on the speech perception task but had no bearing on their segmentation strategy. Following this explanation of the results, the benefit that was gained by the higher vocabulary listeners from having a greater level of vocabulary knowledge may play a role in some other part of the speech recognition process.

There are, however, other possible reasons for the lexical segmentation results found here. The reason for the lack of segmentation pattern differences may lie in the vocabulary scores of the participants used. It could be considered that the range of vocabulary knowledge in this sample was not broad enough to observe any individual differences between higher and lower vocabulary listeners. In the lower vocabulary group, 12 of the 17 listeners had PPVT standard scores of 100 or more. Therefore, although these listeners had lower vocabulary scores relative to the higher vocabulary group, their scores were still equal to or slightly higher than the mean for the general population. Only 5 of the listeners from the lower vocabulary group had PPVT standard scores of less than 100 and all were within the normal range. A greater difference may have been seen between the two vocabulary groups in cue use if there was a greater difference in their receptive vocabulary scores. Choe et al. (2012) were able to compare the segmentation patterns of the 22 best-performing listeners to the 22 worst-performing listeners from a larger group of 88 participants, and so likely had a greater spread of listening ability in their sample for an effect to be seen. A similar approach would be useful in further studies that examine vocabulary and lexical segmentation.

An alternative explanation that accounts for the results found here in the context of the hierarchical model is that the listening conditions presented in this study did not allow a differential segmentation strategy effect to be seen. The conditions may have been degraded in such a way that lexical information was unavailable to all listeners. Even the higher vocabulary listeners needed to use syllabic stress to identify word boundaries. Where there may have been lexical cues available, the syllabic stress cues may have been more numerous or prominent. This explanation seems unlikely, however, given that the large range in listener performance across the four SNR conditions which

would suggest that there were multiple levels of intelligibility and therefore opportunities to utilise multiple levels of the hierarchy of cues.

It is possible that one of the reasons that the hypothesised effect could not be seen was the result of weaknesses in the hierarchical framework that need to be resolved. The findings of the current study, in which the hierarchy did not predict the results found, could be explained in this way. Further evidence for this explanation are the results of Choe et al. (2012) which showed both the better and poorer listeners using stress in the high intelligibility condition, where the Mattys et al. (2005) model would have predicted that listeners would not have used stress, relying instead on other cues from higher levels of the hierarchy. Moreover, in the Choe et al. (2012) study, the better listeners ceased using stress to segment when they encountered the low intelligibility stimuli. This pattern of cue use is inconsistent with the Mattys (2005) model. Therefore, while there is evidence to suggest that, in general, listeners do conform to this hierarchy of segmentation cues depending on the available listening conditions, it appears as if further investigation into the nature of segmentation cue use should be done in order to develop a truly robust and reliable framework for demonstrating how listeners use particular cues in segmentation.

Another potential explanation is that the hierarchical framework is valid, but is not particularly sensitive to individual differences. The framework set out by Mattys et al. (2005) may effectively capture segmentation patterns of listeners as a group, but may not distinguish individual segmentation differences. However the evidence from Choe et al. (2012), showing that better listeners did use stress cues less than poorer listeners, and Sanders et al. (2002) where non-native speakers of English with greater knowledge of the language used more lexical cues than speakers with nearly no knowledge of English, is inconsistent with this account of the results.

One further reason for the findings is that higher vocabulary knowledge does not in fact cause better speech recognition performance. The hypotheses of this study regarding lexical segmentation were based on the assumption that the higher vocabulary listeners' superior lexical knowledge was the reason for their better performance. However, it may have been the case that, while there was a correlation between vocabulary score and speech recognition performance, the two are not causally

related. Accordingly, it is logical that there is no difference in cue use to be seen between groups once they had been divided based on vocabulary score. If the better listeners' speech recognition ability is not a result of their larger vocabularies, then other reasons must be sought. For example, it may be the case that the reason these listeners perform better is that they are more intelligent. Because intelligence is correlated with vocabulary knowledge (Bell, Lassiter, Matthews & Hutchinson, 2001), the better listeners also have higher vocabulary scores. Further investigation into reasons why some listeners perform better than others may provide some explanation for why higher vocabulary listeners performed better.

### **4.3 Word Substitution Errors**

Listeners from both vocabulary groups exhibited similar word substitution error patterns. The higher vocabulary listeners did not produce any greater proportion of word substitution errors that were phonemically similar to the target, as had been previously expected. Hence, it appeared that at least for the current analysis, superior lexical knowledge did not assist listeners in the process of phonemic mapping. That is, vocabulary did not help listeners match the incoming acoustic-phonetic information to stored lexical representations once segmentation had occurred or help them make more accurate attempts at the target word. The lack of differences in WSEs between the higher and lower vocabulary listeners also suggest that having a larger vocabulary did not aid listeners in resolving instances of phonemic ambiguity.

These results were somewhat inconsistent with prior studies. Spitzer et al. (2000) found mixed results with regards to phonemic similarity of the WSEs. Both the listeners who transcribed hypokinetic dysarthric speech and those who transcribed ataxic dysarthric speech who had prior familiarisation with the stimuli performed better than control groups. Interestingly, only the familiarised group who transcribed the ataxic speech showed WSEs that were more phonemically similar to the target in comparison to WSEs made by the control group, whereas the familiarised group transcribing the hypokinetic signal did not. Choe et al. (2012) also examined WSEs and found that better listeners were more likely to produce WSEs that contained some of the correct phonemes in comparison to the poorer listeners. Choe et al. (2012) used mixed dysarthric speech, secondary to

amyotrophic lateral sclerosis, as the stimuli. It is possible that the degree of phonemic similarity to the target depends in part on the type of stimuli used. In Spitzer et al. (2000) and Choe et al. (2012) different types of dysarthric speech, which degrade the speech signal in different ways, were used as stimuli and dissimilar results were found with each. In the current study, the signal was degraded using multitalker babble. Therefore, it may be the case that differences in proportions of WSEs with phonemic similarity can be attributed to different stimuli. In order to make a true comparison, more research using similar types of stimuli as were used in the current study need to be used.

The WSEs that were judged to be phonemically similar to the target words comprised of errors with either consonants correct or vowels correct. There were no differences between higher vocabulary and lower vocabulary listeners in the proportion of WSEs with consonants correct or in the proportion of WSEs with vowels correct. This means that neither group attended more to acoustic-phonetic cues that would identify the consonants. Conversely, neither group attended to cues that assisted in the identification of vowels more than the other.

In both groups, the largest proportion of WSEs had no phonemic resemblance to the target. This finding may reflect that these experimental conditions do not provide listeners with much opportunity to produce phonemically similar word substitution errors. The listening conditions may have been too degraded for listeners to extract phonemic information from the speech stream accurately. In future, the experimental conditions could include more favourable SNR conditions so that more phonemically similar WSEs are produced, which would allow for further comparison between errors produced by higher and lower vocabulary listeners.

The lack of differences between the WSEs produced by the two vocabulary groups may reflect that there are no differences in their approach to phonemic mapping or, as Spitzer et al. (2000) suggest, it may reflect inadequacies in this level of analysis. In future, research that involves a more fine-grained approach may uncover more differences. Alternatively, retaining the same method of analysis but using a much larger dataset may allow more differences to come to light.

Like the LBE results, the lack of phonemic differences between the higher and lower vocabulary groups in the WSEs means that this analysis has not provided answers to the question of why the higher vocabulary group performed better on the original speech perception task.

#### **4.4 Number of Attempts**

The higher and lower vocabulary groups did not differ in the number of attempts they made at responding to the stimulus phrases, both in terms of the average proportion of phrases to which listeners responded, and the average number of words produced in each response. This finding was predicted by the study's hypothesis. Furthermore, both groups of listeners produced similar numbers of LBEs and WSEs. These results demonstrate that better speech recognition performance is not due to greater effort or a larger number of response attempts. Therefore, as the relationship between vocabulary and speech recognition performance in this case cannot be explained by the higher vocabulary listeners simply trying harder to respond to the target, an alternative explanation is needed in order to account for why higher vocabulary listeners displayed better speech recognition ability.

#### **4.5 Limitations and Future Research**

A significant limitation of the study is that the number of errors produced by the participants of this study was small. Consequently, the final dataset used for analysis was smaller than those of other studies examining the same types of errors. For example, McAuliffe et al. (2013) had a similar number of participants. The total number of lexical boundary errors analysed in that study was 670, whereas the total number of lexical boundary errors in the current study was 141. Even though the participants in the current study only heard half the number of target phrases as the McAuliffe et al. (2013) listeners, the number of LBEs produced is nevertheless comparatively very small. Regarding word substitution errors, Spitzer et al. (2000) had the same number of participants who produced over 700 word substitution errors. In the current study, the participants made a total of 504 word substitution errors. There are a number of possible reasons for the low number of errors in the dataset.

It may be that the range of SNR conditions was not ideal for the production of lexical boundary or word substitution errors. In the poorest SNR, -8, many of the listeners did not respond at

all as this SNR made the target phrase too difficult to recognise. On the other hand, at the best SNR, 4, there were many instances where listeners perceived the entire phrase correctly. Therefore in these two SNR conditions, there was little opportunity for listeners to make lexical segmentation or word substitution errors. A solution to this issue would be to decrease the range of the SNR conditions, so that the least favourable condition has a slightly better SNR and the most favourable condition has a slightly worse SNR. Dalrymple-Alford (2014) found that vocabulary knowledge was a greater predictor of performance at the mid-range SNR conditions, so, in any case, focusing on these mid-range conditions may prove more beneficial in research aiming to explore effects of vocabulary on speech perception.

Another reason for the low number of errors in the dataset may be that the listeners became discouraged by the level of degradation of the signal and so did not provide as many responses as they could have. Having a greater number of responses would have elicited a larger number of errors for analysis. Although the listeners were encouraged to guess when attempting to repeat what they heard, it may have been useful to further reiterate the importance of making an attempt even when the noise level is high.

Additionally, in future more participants should be included in the data collection process in order to ensure that a sufficient number of errors are produced for analysis.

A further limitation of the study, as previously mentioned, was the small range of vocabulary scores among the group. This limitation could be addressed in future research by including a larger number of participants. Furthermore, the majority of the participants in this study were university students, who would presumably be more likely to have greater vocabulary knowledge than age-matched peers who did not have university education. Therefore, in order to increase the range of vocabulary scores, and also make the results more applicable to the general population, listeners without university education should be included in the participants.

Future research should focus on more extensive analysis of cue use in speech perception. This study aimed to show whether there was a difference in the extent that listeners used syllabic stress in

segmentation and predicted that if one group of listeners were using stress to a lesser extent it would indirectly show that those listeners were using lexical cues instead. Therefore, any differences between groups of listeners in lexical cue use are by inference only. Future research should assess the use of lexical cues more directly. Studies which compare the use of sublexical cues and lexical cues, and investigate which cues listeners tend to use more when presented with multiple cues, are particularly useful. Existing examples of such studies include Mattys et al. (2005), Sanders & Neville (2000), and Newman et al. (2011). More research needs to be done investigating the extent to which lexical cues, as opposed to sublexical cues, play a role in segmentation. In particular, the Mattys et al. (2005) model of segmentation needs further revision, with the support of more studies that compare the relative importance of particular cues in the process of segmentation.

This study found that listeners attend to the same types of cues when segmenting speech, regardless of their level of vocabulary knowledge. If, as these findings suggest, the reason for higher vocabulary listeners achieving greater success on speech recognition tasks cannot be attributed in any part to differences in lexical segmentation, acoustic-phonetic mapping, or number of response attempts, then further research needs to be conducted that examines other possible explanations. In the original study (Dalrymple-Alford, 2014) working memory ability was assessed using the WAIS-IV (2007) subtest. It was found that neither working memory nor age influenced speech recognition accuracy. The reason for higher vocabulary listeners performing better appears to be difficult to resolve. Further research replicating this study on a larger scale should be done to assess whether the effect of vocabulary on performance can be repeated, and whether any further differences can be found between better and poorer listeners in terms of speech perception strategy usage or outcomes on other assessment tasks.

## **4.6 Conclusion**

The outcomes of this study indicated that a listener's vocabulary knowledge had no differential bearing on the cues used for lexical segmentation, nor did it have an effect on the type of word substitution errors made. The findings of this study suggest that lexical segmentation strategy cannot explain why listeners with greater vocabulary knowledge perform better on a speech



perception in noise task. However, there were considerable limitations that restrict the generalizability of the findings made here. Further research needs to be done in order to ascertain why higher vocabulary listeners perform better on speech perception tasks in adverse conditions.

## References

- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, 47(s2), S53-S71.
- Alamsaputra, D. M., Kohnert, K., Munson, B., & Reichle, J. (2006). Synthesized speech intelligibility among native speakers and non-native speakers of English. *Augmentative and Alternative Communication*, 22(4), 258-268.
- Başkent, D. (2010). Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores. *The Journal of the Acoustical Society of America*, 128(4), EL169-EL174.
- Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the Peabody Picture Vocabulary Test - Third edition and Wechsler Adult Intelligence Scale - Third edition with University students. *Journal of clinical psychology*, 57(3), 417-422.
- Benard, M. R., Mensink, S. J., & Başkent, D. (2014). Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. *The Journal of the Acoustical Society of America*, 135(2), EL88-EL94.
- Benichov, J. C., L.C.; Tun, P.A.; Wingfield, A. . (2012). Word Recognition Within a Linguistic Context: Effects of Age, Hearing Acuity, Verbal Ability, and Cognitive Function. *Ear & Hearing*, 33(2), 250-256.
- Bonding, P. (1979). Frequency Selectivity and Speech Discrimination in Sensorineural Hearing Loss. *Scandinavian Audiology*, 8(4), 205-215.
- Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual Learning of Dysarthric Speech: A Review of Experimental Studies. *Journal of speech, language, and hearing research : JSLHR*, 55(1), 290-305.
- Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61(1-2), 93-125.

- Choe, Y., Liss, J. M., Azuma, T., & Mathy, P. (2012). Evidence of cue use and performance differences in deciphering dysarthric speech. *The Journal of the Acoustical Society of America*, 131(2), 112-118.
- Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to Segment Speech Using Multiple Cues: A Connectionist Model. *Language and Cognitive Processes*, 13(2-3), 221-268.
- Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: the role of prosody. *Journal of Phonetics*, 31(3-4), 585-598.
- Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition*, 25(1-2), 53-69.
- Clopper, C. G., & Bradlow, A. R. (2008). Perception of Dialect Variation in Noise: Intelligibility and Classification. *Language and Speech*, 51, 175-198.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, 119(3), 1562-1573.
- Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, 120(5), 2421-2424.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31(2), 218-236.
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2(3-4), 133-142.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113-121.
- Dalrymple-Alford, J. (2014). *Does Vocabulary Knowledge Influence Speech Recognition in Adverse Listening Conditions?* (Unpublished master's thesis), University of Canterbury, Christchurch, NZ.
- Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59(3), 294-311.

- Divenyi, P. L., Stark, P. B., & Haupt, K. M. (2005). Decline of speech understanding and auditory thresholds in the elderly. *The Journal of the Acoustical Society of America*, 118(2), 1089-1100.
- Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Lisse: Harcourt Test Publishers.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 : Peabody picture vocabulary test*. Minneapolis, MN.: Pearson Assessments.
- Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, 71(6), 1360-1374.
- Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, Reliability, and Validity of PRESTO: A New High-Variability Sentence Recognition Test. *Journal of the American Academy of Audiology*, 24(1), 26-36.
- Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research*, 40(2), 423-431.
- Gow, D. W., Jr., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21(2), 344-359.
- Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech & Language*, 3(4), 367-382.
- Heatley, A. N., I.S.P; Coxhead, A. (2002). Range. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Howes, D. (1957). On the Relation between the Intelligibility and Frequency of Occurrence of English Words. *The Journal of the Acoustical Society of America*, 29(2), 296-305.
- Humes, L. E. (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *The Journal of the Acoustical Society of America*, 112(3), 1112-1132.
- Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology*, 65(8), 1563-1585.

- Jerger, J. J., S.; Pirozzolo, F. (1991). Correlational analysis of speech audiometric scores, hearing loss, age, and cognitive abilities in the elderly. *Ear & Hearing, 12*(2), 103-109.
- Juszyk, P. W. L., P.A. (2002). Speech Perception and Spoken Word Recognition: Past and Present. *Ear & Hearing, 23*(1), 2-40.
- Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America, 122*(1), 418-435.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(01), 1-38.
- Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104*(4), 2457-2466.
- Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. W. (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 107*(6), 3415-3424.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and hearing, 19*(1), 1-36.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1-2), 71-102.
- Mathworks. (n.d.). MATLAB. Natick, MA.
- Mattys, S. L. (2004). Stress Versus Coarticulation: Toward an Integrated Approach to Explicit Speech Segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 397-408.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953-978.
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General, 134*(4), 477-500.

- McAuliffe, M. J., Gibson, E. M. R., Kerr, S. E., Anderson, T., & LaShell, P. J. (2013). Vocabulary influences older and younger listeners' processing of dysarthric speech. *The Journal of the Acoustical Society of America*, 134(2), 1358-1368.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1-86.
- McQueen, J. M. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language*, 39(1), 21-46.
- McQueen, J. M., & Cutler, A. (2001). Spoken word access processes: An introduction. *Language and Cognitive Processes*, 16(5-6), 469-490.
- Miller, G. A., & Licklider, J. C. R. (1950). The Intelligibility of Interrupted Speech. *The Journal of the Acoustical Society of America*, 22(2), 167-173.
- Monaghan, P., White, L., & Merkx, M. M. (2013). Disambiguating durational cues for speech segmentation. *The Journal of the Acoustical Society of America*, 134(1), EL45-EL51.
- Munson, B. (2001). Relationships Between Vocabulary Size and Spoken Word Recognition in Children Aged 3 to 7. *Contemporary Issues in Communication Science and Disorders*, 28, 20-29.
- Nation, I. S. P., & Heatley, A. . (1994). Range: A program for the analysis of vocabulary in texts. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Newman, R. S., Sawusch, J. R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech. *Journal of Memory and Language*, 64(4), 460-476.
- Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1209-1228.
- Orr, D. B., Friedman, H. L., & Williams, J. C. C. (1965). Trainability of listening comprehension of speeded discourse. *Journal of Educational Psychology*, 56(3), 148-156.
- Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology*, 47(s2), S72-S82.

- Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *The Journal of the Acoustical Society of America*, 94(4), 2027-2035.
- Rubenstein, H., & Pollack, I. (1963). Word predictability and intelligibility. *Journal of Verbal Learning and Verbal Behavior*, 2(2), 147-158.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, 35(4), 606-621.
- Sanders, L. D., & Neville, H. J. (2000). Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research*, 43(6), 1301-1321.
- Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech Segmentation by Native and Non-Native Speakers: The Use of Lexical, Syntactic, and Stress-Pattern Cues. *Journal of Speech, Language & Hearing Research*, 45(3), 519-530.
- Schneider, B. A., Li, L., & Daneman, M. (2007). How Competing Speech Interferes with Speech Comprehension in Everyday Listening Situations. *Journal of the American Academy of Audiology*, 18(7), 559-572.
- Shafiro, V., Sheft, S., Risley, R., & Gygi, B. (2013). Intelligibility of interrupted speech in listeners of different age and hearing status. *Proceedings of Meetings on Acoustics*, 19(1), 1-7.
- Smith, M. R. C., A.; Butterfield, S.; Nimmo-Smith, Ian. (1989). The Perception of Rhythm and Word Boundaries in Noise-Masked Speech. *Journal of Speech and Hearing Research*, 32, 912-920.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *The Journal of Neuroscience*, 32(25), 8443-8453.
- Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *The Journal of the Acoustical Society of America*, 122(6), 3678-3687.
- Spitzer, S. M. L., J.M; Caviness, J.N.; Adler, C. (2000). An Exploration of Familiarization Effects in the Perception of Hypokinetic and Ataxic Dysarthric Speech. *Journal of Medical Speech-Language Pathology*, 8(4), 285-293.
- Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some Factors Underlying Individual Differences in Speech Recognition on PRESTO: A First Report. *Journal of the American Academy of Audiology*, 24(7), 616-634.

- van der Lugt, A. H. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics*, 63(5), 811-823.
- Viemeister, N. F., & Wakefield, G. H. (1991). Temporal integration and multiple looks. *The Journal of the Acoustical Society of America*, 90(2), 858-865.
- Vitevitch, M. S., & Luce, P. A. (1998). When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, 9(4), 325-329.
- Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1), 98-108.
- Wang, X., & Humes, L. E. (2010). Factors influencing recognition of interrupted speech. *The Journal of the Acoustical Society of America*, 128(4), 2100-2111.
- Warren, R. M. (1970). Perceptual Restoration of Missing Speech Sounds. *Science*, 167(3917), 392-393.
- Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2010). Colliding cues in word segmentation: The role of cue strength and general cognitive processes. *Language and Cognitive Processes*, 25(3), 402-422.
- Wechsler, D. (1981). *WAIS-R: Wechsler Adult Intelligence Scale*. New York, N.Y.: Psychological Corp.
- Wechsler, D. (1997). *WAIS III: Wechsler Adult Intelligence Scale*. New York, N.Y.: Psychological Corp.
- Wechsler, D. (2001). *WTAR : Wechsler test of adult reading*. San Antonio, Tex.: Psychological Corp.
- Wechsler, D. (2008). *WAIS-IV : Wechsler adult intelligence scale*. San Antonio, Tex.: Pearson.
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention. *The Journal of Neuroscience*, 32(40), 14010-14021.
- Woodfield, A., & Akeroyd, M. A. (2010). The role of segmentation difficulties in speech-in-speech understanding in older and hearing-impaired adults. *The Journal of the Acoustical Society of America*, 128(1), EL26-EL31.